Uncontrolled mutation of super-intelligent machines may lead to the destruction of humanity

Deep Bhattacharjee [1] and Sanjeevan Singha Roy [2]

[1]AATWRI – R&D Directorate of Electro-Gravitation Simulation & Propulsion Laboratory
[2]Affiliation not available

October 30, 2023

Abstract

If in future, the highly intelligent machines control the world, then what would be its advantages and disadvantages? Will, those artificial intelligence powered superintelligent machines become an anathema for humanity or will they ease out the human works by guiding humans in complicated tasks, thereby extending a helping hand to the human works making them comfortable. Recent studies in theoretical computer science especially artificial intelligence predicted something called 'technological singularity' or the 'intelligent explosion' and if this happens then there can be a further stage as transfused machinery intelligence and actual intelligence where the machines being immensely powerful with a cognitive capacity more than that of humans for solving 'immensely complicated tasks' can takeover the humans and even the machines by more intelligent machines of superhuman intelligence. Therefore, it is troublesome and worry-full to think that 'if in case the machines turned out against humans for their optimal domination in this planet'. Can humans have any chances to avoid them by bypassing the inevitable 'hard singularity' through a set of 'soft singularity'. This paper discusses all the facts in details along with significant calculations showing humanity, how to avoid the hard singularity when the progress of intelligence is inevitable.

Uncontrolled mutation of super-intelligent machines may lead to the destruction of humanity

Deep Bhattacharjee[1]*, **Sanjeevan Singha Roy**[2]
[1]*Departmental In-Charge of AATWRI – R&D Directorate of Electro-Gravitation Simulation & Propulsion Laboratory, Bhubeneshwar, Odisha, India*
[1]*Research Mentor in Research Convention, Chandigarh, Punjab, India*
[2]*Department of Physics, Birla Institute of Technology, Mesra, Jharkhand, India*
(28th June 2021)

Abstract – If in future, the highly intelligent machines control the world, then what would be its advantages and disadvantages? Will, those artificial intelligence powered superintelligent machines become an anathema for humanity or will they ease out the human works by guiding humans in complicated tasks, thereby extending a helping hand to the human works making them comfortable. Recent studies in theoretical computer science especially artificial intelligence predicted something called 'technological singularity' or the 'intelligent explosion' and if this happens then there can be a further stage as transfused machinery intelligence and actual intelligence where the machines being immensely powerful with a cognitive capacity more than that of humans for solving 'immensely complicated tasks' can takeover the humans and even the machines by more intelligent machines of superhuman intelligence. Therefore, it is troublesome and worry-full to think that 'if in case the machines turned out against humans for their optimal domination in this planet'. Can humans have any chances to avoid them by bypassing the inevitable 'hard singularity' through a set of 'soft singularity'. This paper discusses all the facts in details along with significant calculations showing humanity, how to avoid the hard singularity when the progress of intelligence is inevitable.

Keywords – ARTIFICIAL Intelligence; TECHNOLOGICAL Singularity; INTELLIGENT Explosion; SUPER-HUMAN Level Machine Intelligence; SOFT Singularity.

Introduction – Human life is balancing on the edge of machines. The edge is like a halo point. Too much dependency on the edge will destroy the balance which leads to unavoidable circumstances through which humans may loose all their controls on machines and those machines will rise in their utmost potentials through logical and analytical perspectives with some intelligent parameters which are sufficient enough to bring a harm or even destruction on the entire human life's, who has been worshipping machines until now and also through some points in future for the ease of their works. Superintelligent machines capable of being superintelligence may not come in one day and may not come in the near future that we can easily imagine. But, it is evident and based on the progression of intelligent software's embedded in smart hardwires, an expulsion rather an explosion of machine intelligence is inevitable. However, one thing that we didn't care about is that we have already initiated the process of making a super-intelligent machine. That is, although weakly coupled to human level intelligence comparison, we have succeed in making smart artificial intelligent (AI) machines that smarten our culture and works making us more dependable on them for the sake of our own comforts. Those AI machines, in future first transformed to human level machine intelligence (HLMI), then to super-human level machine intelligence (SHLMI), crafting the way for a 'technological singularity' (TS) where machines will overrun human intelligence (HI), thus machines will reproduce to much smarter machines with ultra-smart programs equipped with ultra-smart software's and hardwires, that in essence could transform the SHLMI and TS to Actual intelligence (AcI) which in turn makes the pathway for a domination of humans, or in other way, just like we dominate the chimps by caging them in zoos although we have evolved from them, in the same way, those AcI machines will dominate and caged humans with their ultimate form of powers making humans as slaves which in course of time leads to human destruction or human extinction, when machines will rule this world. Those AcI machines will be dependable on some free parameters which they will try to develop to their fullest capacity leading to a more intelligent form of transfused machines (TM) that will outbreak all the laws of machines thereby destroying machines and takeover earth from them through their transfused actual intelligence (TAcI) that will lead to a further development that is beyond the scope of imagination standing the in present day scenario of the infancy of AI in the machines. Therefore, humans must have to find a way to balance and keep balancing on the edge although the edge is getting sharper and sharper as time proceeds towards future. This inclusion of humans and sharper edges aka., machine intelligence will lead to a technique that will be explained in details in the paper called the error approximated machine intelligence (EAMI) that transformed humans by creating a domain wall intelligence (DWI) which in particular leads to a more perceivable soft singularity (SS) before entering and warning a hard singularity (HS) or

not even entering the HS but bypassing HS through some parametric cancellations that enforce humans to keep the MI dependable on humans preventing the human destruction by those AcI machines keeping both the machines and humans safe from the evils of SI, SHLMI, AcI, and machine dominated TMs.

Approaches – Time is not just an one way absolute entity [1]. Rather it's a flexible and relativistic [2], realistic norm that has been specifically destined for the improvement purposes with the improvement norm I. Developing with regards, to a function of time as, $\frac{\partial I}{\partial t}$ it gives a satisfactory result $\dot{I}$. In the computer science, theoretically time itself is a perception upon which there is a spatial dependency, as to say, space shrinks with time. More, intelligent machines are being developed in course of time taking less space than usual [3], but that does not qualify for being a TMAcI machine or HLMI machine to be of very less spatially occupying, maybe the forever notion of $\dot{I}$ rebounds in this cases as $-\dot{I}$ that regards for a more modified development over the course of time. The temporal and spatial dependency parameter can be chalked out as $t + s$ that when averaged out over the course of improvement then gives a relation as $\dot{I} = \langle \sigma\tau \rangle$, where $\sigma = \prod_{a\,=\,HLMI}^{TMAcI} s_a$ and $\tau = \prod_{a\,=\,HLMI}^{TMAcI} t_a$ qualifies for a squared norm value of,

$$\left|\dot{I}\right|^2 \equiv \prod\nolimits_{a\,=\,HLMI}^{TMAcI} s_a + \prod\nolimits_{a\,=\,HLMI}^{TMAcI} t_a \;\sim 0$$

It is indeed important to note that $\dot{I}$ in $\left|\dot{I}\right|^2$ is indeed positive definite because of its dependency on 5 roots as the parametric stabilization regards to $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ taken the values to be of the respective levels or orders,

$$\left.\begin{matrix} AI, & +\theta_1 \\ HLMI, & +\theta_2 \\ SHLMI, & +\theta_3 \\ AcI, & +\theta_4 \\ TMAcI, & +\theta_5 \end{matrix}\right\} + \; variables$$

$$\left.\begin{matrix} AI, & -\theta_1 \\ HLMI, & -\theta_2 \\ SHLMI, & -\theta_3 \\ AcI, & -\theta_4 \\ TMAcI, & -\theta_5 \end{matrix}\right\} - \; variables$$

The inert difference between the + and – variables with regards to the roots of θ_i can be presented as + being beneficial to humankind, − being destructive to humankind, the question that asserts from here is that will destruction inevitable having the parameter value $-\gg +$ over the period of $\dot{I}$. To answer this question, it is necessary to introduce the norm of the matrices that belongs to each of the permutation groups of the roots of θ_i, the norm being taken on the assumption, is that the matrices δ_{ij} [which corresponds to two values as the matrix element, 0 & 1 the former being singularity, the later being not] can integrate over the average spatial and temporal components from the present notion of the time stamp to the future on the boundary value of θ_i to denote the entire path of the operations made in FLOPS (or Floating Points Operations Per Second),

$$\int_{\partial\theta_i} Det\;\delta_{ij} \approx FLOPS(10^x), \quad with\; x \gg 0\; and\; x \in \mathbb{R}$$

As we see that, the computation powers will increase with time, so if the norm $\left|\dot{I}\right|^2$ satisfies the relation $-\gg +$ then, we have to force that there must have some intrinsic calculations going on with – that in essence is the intelligent parameters which is also the root of the intelligent variables Ψ^2 giving the relation as, $-\Psi$ couples with $'-'$ and $+\Psi$ couples with $'+'$ to yield the value of the norm,

$$\left|\dot{I}\right|^2_{(\pm\Psi)} \equiv 0$$

This in essence could be termed as the 'singularity' which is at sense, the 'technological explosion' attainable through 'TS' at point 0. The basic question that arises here is that how will the chain of development takes place? And are we bound to attain the norm! or singularity? If so, then how can we bypass the SS by means of creating a DWI through a mechanism of EAMI that in essence be fruitful for not only the human dominating SHLMI but also the machine dominating TMAcI, thereby making the future safe for the humanity. However, to satisfy the intrinsic condition, we also have to be careful so that the integral of the FLOPS remains stable and within the boundary of θ_i. The value of x in 10^x might be large enough for the present mankind to comprehend but, it might not be large enough for the crossing of $\partial\theta_i$ aka., the boundary value of the roots of θ_i to prevent the machine domination over humankind in the faraway future, whose roots lie in the present world. There is a saying that "*fate is inevitable, and one can't escape the fate*", however, its impossible to tell whether fate can be bypassed or not, but as portrayed in LUCY [4] and I, Robot [5], fate really cannot be escaped. But, humans are not so fool and can't handle everything by its fate rather they choose to change the bad fate and favors' it for a good fate. As per the saying of Alan Turing [6], AI can

be attainable by virtue of making a 'child AI' which then accustomed with its surroundings and transformed itself into a more 'matured AI'. Different scientists have come upon a conclusion that machine-brain interference or mimicking neurons of human brains and crafting them in the machines may lead us to a way of achieving HLMI, but the reality is in fact, hitherto, to the approximation of the Turin's word that a child AI might be the protagonist for a matured AI if we can implement the algorithm of thinking in the mind of a AI machines. Then, that AI will evolve in course of time and developed its thinking's saturated with the human intelligence for drafting a HLMI machine. But what are the odds that lies in creating or evolving to a HLMI machines and then SHLMI machines and how its progression can be linked with the scale of human civilization. The benchmark of intelligence limit or advancement limit of human civilization can be best described by the Sagan-Kardashev scales which notes the present scale being 0.73 [7, 8], the attributed formula with P being power is,

$$K = \frac{\log_{10} P - 6}{10}$$

This extrapolates to 3 Types of civilizations (or scales) being attributed as Type I (producing 10^{16} Watts of power), Type II (producing 10^{26} Watts of power), Type III (producing 10^{36} Watts of power) with the current 0.73 having an approximated power consumption of 10 terawatts as of 1970's. Time dependency factor has always been there as regards to the development of human civilizations with regards to the processing powers, the basic being cornered with the Moore's law that '*power of computer chips doubled every two years*' [9]. This statement holds true as long as we are in the attainment scale of HLMI but the development period shortens to much less than 2 years for the attainment of SHLMI from HLMI and much much less for AcI from SHLMI and thus so far. Thus, if we take the improvement dependency with time $\dot{I}$, the roots can be attributed like this fashion,

$$\dot{I}_{\theta_i} = Transition\ Time: \begin{matrix} \pm\theta_1 & Span^1 \\ \pm\theta_2 & Span^2 \\ \pm\theta_3 & Span^3 \\ \pm\theta_4 & Span^4 \\ \pm\theta_5 & Span^5 \end{matrix}$$

With the time transition from each successive span to the other being decreasing or rapid improvements like $Span^5 < Span^4 < Span^3 < Span^2 < Span^1$. This declination implies the machines intelligence to evolve from one span to another taking much less time than the preceding spans. The important point is that each roots of θ_i contains two factors '$+$' and '$-$' with the former being favorable to humans while the later being harmful for humans. It can be said with a higher degree of precession that from $Span^2$ onwards the '$-$' sign will dominates '*and that is very much natural*' unless we do something intelligent like EAMI to create a DWI for hitting the SS by bypassing the HS. In that case, the formal question arises as to: if the HS is destined to attain at some point around $\pm\theta_3$ then, if we bypass HS to SS around $\pm\theta_3$ *will* $\pm\theta_4$ *and* $\pm\theta_5$ *comes into existence?* Well, humanity is not sure about that, and there lies a probability which can be expressed as,

$$\pm\theta_1 \longrightarrow \pm\theta_2 \longrightarrow \pm\theta_3 \ ? \ \pm\theta_4 \ ? \ \pm\theta_5$$

Right now, AI is still in its infancy. It is on the verge of development. Different AI devices and robots are being developed across various parts of the world with the population of AI robots exceeds 10 million worldwide. But, the fact to consider is that the principles of AI are not *bundled* rather they are *segregated*. The robots that play chess doesn't have the capacity to perform cleaning, and the robots that is tasked to cleaning couldn't perform complicated surgeries. Therefore, proper bundled parameters of algorithm must be encapsulated in the AI robots such that, a single robot can perform many tasks at a more better and intelligent way than that of humans. Humans being held as the sole programming authority to AI and the way the AI is prosperising, Turin's statement about a protagonist child AI bundled with several functions is not too far to achieve. And if this can be achieved then humans through some clever computation can implement the *classifiers* of that child AI for choosing or harnessing the better aspects of technology from its ambient aspects to implement in them for attaining the HLMI. However, to develop HLMI it needs a good deal of processing powers and automation to generate the goal of creating *a cognition responsible for complex tasks like humans*. HLMI can't be solely manufactured. It needs to be evolved. And that evolution takes the classifiers to classify the best among best and worst and adopting with it in the same way just as humans have evolved over the period of time from apes to modern day intelligent mans. There lies an intrinsic shortcut for the AI to achieve HLMI unlike the case of humans which took more than 1,00,000 years from caveman to skylinemans. AI being equipped with the sophisticated algorithms inputted by the humans can easily evolve into HLMI with a prospect for further development in a course of a much shorter time than the previous evolution from AI to HLMI. HLMI once gets successfully crafted in our society, their intelligence will be same as that of humans and they in turn can make themselves

more modified with *a more complex super-human cognitive capabilities* for developing SHLMI. It is regardless to say that SHLMI are regardless to the aspect of *human dominated AI*. If humans let them to evolve further then at some point they will turn either useful or harmful for the humans, with the axis being inclines for being the more harmful parameters. Therefore, at the onset of HLMI, humans have to inject bruteforcely the code of EAMI which can be classified as (the symbolic cancellation operator) we see the formula,

$$\otimes \frac{Det\ \delta_{ij} \times \beta}{\alpha}, \qquad \exists\ \beta, \alpha = Negative\ value$$

Here, δ_{ij} stands for the matrix of 'i' rows and 'j' column whose determinant denotes the boundary values of the roots as $\partial\theta_i$ with the β parameter being the '*friendly human algorithm being injected in the AI machine*' and α being *the degree of the algorithm to be injected*. Therefore, for convention the matrix δ_{ij} can take 2 values as,

- $\left(\delta_{ij} = \begin{matrix} 0 & 1 \\ 1 & 1 \end{matrix} = Det - 1\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 1 \\ 1 & 0 \end{matrix} = Det - 1\right)$
- $\left(\delta_{ij} = \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} = Det - 1\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 0 & 1 \\ 0 & 1 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 0 & 0 \\ 0 & 1 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 1 \\ 0 & 0 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 0 & 0 \\ 1 & 1 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 0 \\ 1 & 0 \end{matrix} = Det\ 0\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} = Det\ 1\right)$
- $\left(\delta_{ij} = \begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix} = Det\ 1\right)$

Therefore, the matrix element being 0 and 1, in all permutations it can give the determinant as -1, 0 and 1 where we will ignore the 0. If β is taken as a negative value with α being any negative number such that the large value of α leads to a lesser degrees of freedom, we will find the EAMI parameter if we take Det δ_{ij} as -1, then $\otimes \frac{Det\ \delta_{ij} \times \beta}{\alpha}$ will be negative and this denotes a cancellation of the negative roots of θ_i where if the 'cancellation could be done on $\pm\theta_3$ then, there will be only $+\theta_3$ which will ultimately evolve to an AcI at root $+\theta_4$ containing only the beneficial factor '+' thus avoiding hitting a HS although through the journey 7 *SS* needs to be cleared which acts as a way of bypassing HS in a mimicking SS. Therefore, a DWI can be created around $+\theta_3$ which will transcend to $+\theta_4$ in the way of a beneficial AcI and furthermore into $+\theta_5$ as beneficial TM that regards for an optimal development. The transition from SHLMI to TM or $+\theta_4$ to $+\theta_5$ is difficult and assumed to take place in $K = 3.0$ scale or Type III civilization. In the case of HLMI, humans will make the more modified machines, but from HLMI to SHLMI machines will reproduce to a further developed machines containing developed chips with ultra-smart hardwire and software but, in case of TMs there will be a cocktail of Human-Machine accumulation or humans are evolved to function as machines, or in a word, machine itself doesn't exists in hardwire form, but exists in the form of super-intelligent bio-fluids humans will act as a replica for machines where the difference between humans and machines faded way into a human-machine transfusion, that is not only hard to believe but also, hard to imagine. If EAMI can be properly implemented then the path to superintelligence can be of the form [10],

$$\pm\theta_1 \longrightarrow +\theta_2 \longrightarrow +\theta_3 \longrightarrow +\theta_4 \longrightarrow +\theta_5 \longrightarrow \ldots$$

However, there is another parameter called 1, which we haven't take into account yet. This will determine a backreaction on $\otimes \frac{Det\ \delta_{ij} \times \beta}{\alpha}$ where, due to the positive value of $\frac{Det\ \delta_{ij} \times \beta}{\alpha}$ as Det of δ_{ij} with the cancellation occurring with the positive values of θ, the negative value remains same as $-\theta$ the chain will continue with the destructive superintelligence as follows,

$$\pm\theta_1 \longrightarrow -\theta_2 \longrightarrow -\theta_3 \longrightarrow -\theta_4 \longrightarrow -\theta_5 \longrightarrow \ldots$$

Which indicates that there is scope left for the machines to take over humanity and the destructive superintelligence throughout, that can be negated with a more tough parameter imposed on $-\theta_2$ to break the chain as we will *locally* take the value of α as negative but β to be positive (where humans will inject more errors into the system for cancellation of a destructive AI), then, the formulae leads to an EAMI 'symbolic cancellation parameter' as,

$$(\mathcal{L})\otimes \frac{Det\ \delta_{ij} \times \beta}{\alpha}, \qquad \exists\ \beta = Positive\ value\ \&\ \alpha = Negative\ value$$

Where $\mathcal{L}$ stands as a local function then, even if the value is 1, the machine will never give any back reaction and the chain continues as,

$$\pm\theta_1 \longrightarrow +\theta_2 \longrightarrow +\theta_3 \longrightarrow +\theta_4 \longrightarrow +\theta_5 \longrightarrow \ldots$$

Thus we get a friendly AI machines to reproduce further without any human destruction. One thing that needs to be mentioned is that the DWI will be in between $\pm\theta_2$ and $\pm\theta_3$.

---

*itsdeep@live.com (corresponding author)
‡ Actual intelligence robot AUTIZMO based on CXAI technology (where history meets future) has been invented by Dr. Catherine Demetriades, a prominent particle physicist in CERN, Geneva, Switzerland. Email 1: catatrix@cxaitechnology.com, Email 2: Catherine.Demetriades@CERN.ch
†^{0} Author has no conflicting interests as related to this paper.
†^{1} No source of funding has been assigned for this work.

References –

[1] Milo, Rafi. "Absolute Time and Absolute Simultaneity." *Physics Essays*, vol. 28, no. 1, 2015, pp. 24–31. *Crossref*, doi:10.4006/0836-1398-28.1.24.
[2] Hsu, J. P. "The Analysis of Time: Is the Relativistic Time Unique?" *Foundations of Physics*, vol. 9, no. 1–2, 1979, pp. 55–69. *Crossref*, doi:10.1007/bf00715051.
[3] Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Reprint, Oxford University Press, 2016.
[4] Seitz, Matt Zoller. "Lucy Movie Review & Film Summary (2014) | Roger Ebert." © *Copyright 2021*, 2014, www.rogerebert.com/reviews/lucy-2014.
[5] Bauer, Patricia. "I, Robot | Summary, Characters, & Facts." *Encyclopedia Britannica*, 1998, www.britannica.com/topic/I-Robot.
[6] Agar, Jon. *Turing and the Universal Machine (Icon Science): The Making of the Modern Computer*. Reprint, Icon Books Ltd, 2017.
[7] Kaku, Michio. "The Physics of Interstellar Travel : Official Website of Dr. Michio Kaku." *The Physics of Interstellar Travel*, 2010, mkaku.org/home/articles/the-physics-of-interstellar-travel/#:%7E:text=To%20one%20day%2C%20reach%20the%20stars.&text=Even%20the%20familiar%20stars%20we,interstellar%20travel%20to%20be%20practical.
[8] Sagan, Carl, et al. *The Cosmic Connection: An Extraterrestrial Perspective*. Ishi Press, 2019.
[9] Tardi, Carla. "Moore's Law Explained." *Investopedia*, 2021, www.investopedia.com/terms/m/mooreslaw.asp.
[10] Bhattacharjee, Deep. "The Clampdown Effect: On The Expulsion of Super-Intelligence". *Preprints* 2021, 2021040353, 1–7. https://dx.doi.org/10.20944/preprints202104.0353.v1